EXPERT SYSTEM VERIFICATION AND VALIDATION STUDY

Delivery 3A & 3B - Trip Summaries

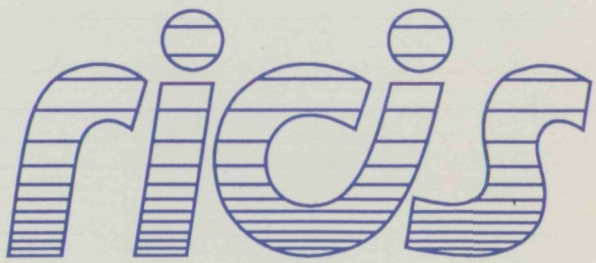
Scott French

International Business Machines Corporation

August 23, 1991

Cooperative Agreement NCC 9-16
Research Activity No. AI.16

NASA Johnson Space Center
Information Systems Directorate
Information Technology Division



Research Institute for Computing and Information Systems
University of Houston-Clear Lake

TRIP REPORT

The RICIS Concept

The University of Houston-Clear Lake established the Research Institute for Computing and Information Systems (RICIS) in 1986 to encourage the NASA Johnson Space Center (JSC) and local industry to actively support research in the computing and information sciences. As part of this endeavor, UHCL proposed a partnership with JSC to jointly define and manage an integrated program of research in advanced data processing technology needed for JSC's main missions, including administrative, engineering and science responsibilities. JSC agreed and entered into a continuing cooperative agreement with UHCL beginning in May 1986, to jointly plan and execute such research through RICIS. Additionally, under Cooperative Agreement NCC 9-16, computing and educational facilities are shared by the two institutions to conduct the research.

The UHCL/RICIS mission is to conduct, coordinate, and disseminate research and professional level education in computing and information systems to serve the needs of the government, industry, community and academia. RICIS combines resources of UHCL and its gateway affiliates to research and develop materials, prototypes and publications on topics of mutual interest to its sponsors and researchers. Within UHCL, the mission is being implemented through interdisciplinary involvement of faculty and students from each of the four schools: Business and Public Administration, Education, Human Sciences and Humanities, and Natural and Applied Sciences. RICIS also collaborates with industry in a companion program. This program is focused on serving the research and advanced development needs of industry.

Moreover, UHCL established relationships with other universities and research organizations, having common research interests, to provide additional sources of expertise to conduct needed research. For example, UHCL has entered into a special partnership with Texas A&M University to help oversee RICIS research and education programs, while other research organizations are involved via the "gateway" concept.

A major role of RICIS then is to find the best match of sponsors, researchers and research objectives to advance knowledge in the computing and information sciences. RICIS, working jointly with its sponsors, advises on research needs, recommends principals for conducting the research, provides technical and administrative support to coordinate the research and integrates technical results into the goals of UHCL, NASA/JSC and industry.

EXPERT SYSTEM VERIFICATION AND VALIDATION STUDY

Delivery 3A & 3B - Trip Summaries

ES V&V Guidelines/Workshop Task
RICIS Contract #69
Deliverable #3 - Trip Summary

August 23, 1991

3700 Bay Area Blvd. Houston, Texas 77508-1199

Preface

This document contains information pertaining to deliverable #3 as specified in RICIS contract #69. The information contained in this document pertains to the AAAI Verification, Validation and Testing Workshop held in Anaheim, CA during July of 1991. Scott French of IBM attended the conference described herein and has prepared the summary that follows.

Workshop on Knowledge-Based Systems Verification, Validation and Testing at the 9th National Conference on Artificial Intelligence

This appendix serves to document key results from attending the fourth workshop on Verification, Validation and Testing (VV&T) held on July 17, 1991 as part of the 9th national conference on AI.

The most interesting part of the workshop was the first part. Representatives from the U.S., Japan and Europe presented surveys of VV&T within their respective regions. The first to present was a team from IBM (FSD Houston) and NASA. That presentation focused on the results of a survey done the previous year. This survey attempted to analyze how well VV&T practices are being applied to current ES projects. The final result was a set of recommendations for where future work should focus in the application of VV&T to ES development practice. The representative from Japan spoke to an effort being sponsored by the Japan Information Processing Development Center to make VV&T the *state-of-the-practice*. Their work has focused on defining guidelines that would help ES developers determine what tasks are required to be done in order for that system to be considered *verified*. These guidelines have been captured as a large checklist of VV&T tasks. The checklist also gives some guidance with respect to deciding which techniques apply for the problem at hand. The representatives from Europe (one from Spain and one from France) provided a different perspective on approaching VV&T of ES by discussing the ESPRIT (European Strategic Program of Research about Information Technology) project. This project, involving teams from Denmark, Spain, and France, focuses on tools for automating VV&T rather than guidelines/checklists. A generic interface language (VETA) has been developed that can describe knowledge captured in several different kinds of ES shells. ES's written in these shells can then be converted to VETA constructs which are understood by the collection of VV&T tools that are being developed.

Another interesting part of the workshop focused on current efforts to define industry standards for AI and how that might affect approaches to VV&T of ES. The discussion started with presentations from a panel of people involved with the development of these standards. Dr. Lance Miller of SAIC gave an overview of how standards are developed and the different organizations involved. He also presented a general discussion of why standards will be helpful for the VV&T of ES. The panel was composed of 1 representative from the AIAA working group that is addressing standards, 2 representatives from the IEEE working group on standards, and 1 representative from the Army who gave the DoD perspective on the need for standards. This discussion was basically not very controversial with everyone basically agreeing that standards will be beneficial. One key aspect of this that I think is starting to show signs of a *grass roots* movement is the requirement that all inference engines be certified. Everyone was in unanimous agreement that this kind of standard would enforce something recognized by the VV&T community as being essential to performing VV&T on an expert system.

The next part of the workshop focused on VV&T methods. The papers presented were of a wide variety. The most interesting ones focused on applying mathematically techniques to verification of rule-bases and techniques for capturing information relating to the *process* of developing software. With regard to mathematical analysis techniques, two papers were presented expressing rule-based systems as connected graphs and applying existing algorithms (e.g., shortest path) to those graphs as a method for analyzing the rule-base. Another paper presented an idea for analyzing a rule-base by analyzing the antecedent instead of the consequent and then using propositional logic to determine where there may be problems in the rule-base. The other interesting papers focused on the fact that a significant part of developing an ES involves a lot of information and group decisions. The point was made that much effort has focused on helping an individual analyze an ES, but little effort has been made to help groups (e.g., a code inspection team) make the right VV&T decisions. Part of this focused on describing a system that incorporates many techniques such has multi-media for helping people understand the process within which the development of the ES fits.

The final part of the workshop focused primarily on tools. I found this to be the least informative part of the workshop. There was one slightly interesting paper describing a tool developed in Canada called COVER (a University project). COVER is a tool in the same vein as EVA, but (as you might expect) it is better. Based on the presentation, it clearly does many kinds of analysis on a rule-base and apparently does have some industry use (the author listed 3 companies that currently use COVER).

In summary, the workshop was worthwhile. Many important topics were discussed and even more importantly, everyone seemed to be on the same wavelength as far as what are the real issues of VV&T.

ES V&V Guidelines/Workshop Task
RICIS Contract #69
Deliverable #3 - Trip Summary

August 23, 1991

3700 Bay Area Blvd. Houston, Texas 77508-1199

Preface

This document contains information pertaining to deliverable #3 as specified in RICIS contract #69. The information contained in this document pertains only to the Conference on Methodologies, Tools and Standards for Cost-Effective Reliable Software Verification and Validation sponsored by the Electric Power Research Institute (EPRI). Scott French of IBM attended the conferences described herein and has prepared the summary that follows.

Summary of the EPRI Conference on "Methodologies, Tools and Standards for Cost-Effective Reliable Software Verification and Validation," Aug. 7-9, 1991.

The following is intended to outline the major focus of this conference as it applies to the RICIS guidelines task. The conference combined presentation of papers with "breakout" sessions. These "breakout" sessions (there were 5 different groups) were intended to provide a forum for discussion of the following issues:

- Development process
- Automated Tools
- Software Reliability
- Methods
- Standards
- Cost\Benefit Considerations

Issues and possible solutions for each of these were discussed within these groups with final results being presented on the final day of the conference. The outline below shows the presentations that were attended/participated in.

Aug. 7

- "A Comparative Evaluation of V&V Procedures for Conventional Software and Expert Systems," F. Saglietti
- "REALM Verification and Validation," ConEdison
- Breakout Sessions

Aug. 8

- "Qualified Software Development Methodologies for Nuclear Class 1E Equipment," Spectrum Technologies, USA
- "Formal Specification and Verification for Critical Systems: Tools, Achievements, and Prospects," John Rushby
- "Software Requirements Specifications for Critical Applications, " Ontario Hydro
- "Verification and Validation of Control System Software," Oak Ridge National Laboratories
- "Formal Specifications for Safety Grade Systems," Argonne National Laboratories

- "Guidelines for the Use of Microcomputer Applications in Safety Related Activities," Texas Utilities
- "In Search of Cost-Effective, Reliable Software," EPRI
- "Methodologies for V&V of Expert Systems as a Function of Component, Criticality and Life-Cycle Phase," Lance Miller
- "A Standardized Approach to V&V to Assist ES Development," Ohio State University
- Breakout Sessions

Aug. 9

- "V&V of Real Time AI Systems Using the Activation Framework," Worcester Polytechnic Institute
- "V&V and Standards," SAIC
- "Practical Requirements for Software Tools to Assist in the Validation and Verification of Hybrid Expert Systems," KARTA Technology
- RICIS Survey Work
- Summary of Breakout Session
- Panel on future challenge of V&V
- Open Discussion

Note that on the final day the RICIS survey work was presented (refer to initial RICIS contract #69 work). It was well received and a copy of the survey paper will be placed in the proceedings for the conference. It was interesting to see the context of the discussion change after the survey results were presented. For example, many people were hot on the idea of rapid prototyping as the key to defining a system (especially so they could show immediate progress to their management). After the survey results were presented these same people were speaking to the risks of having flashy rapid prototypes that their management would want made operational.

The papers presented were marginal in content with respect to the insights gained in the breakout sessions. The information disseminated there will be helpful to the guidelines task. The following are some highlights from those discussions:

- groups represented have a very grave concern with the cost of V&V. Analysis of the cost-effectiveness of formal methods is desperately needed

- guidelines are desperately needed to help focus the V&V task (e.g., some kind of cross reference between formal methods and the part of the development process where they apply)
- formal methods need to be made more practical
- some type of "consumer reports" type rating system is needed when purchasing COTS
- guidelines/techniques are needed for measuring software reliability
- guidelines on metrics and how to apply them
- automated tools are a pipe dream ("how can we automate methods when we don't understand how to manually implement those methods")
- guidelines for quantifying cost benefit for presentation to managers
- regulation of the software industry
- improved education in the application of formal methods/processes
- guidelines for definition of acceptance criteria
- classification of software is imperative

The makeup of the conference was primarily engineers and managers. It was clear (this fact was stated many times) that, when it comes to applying formal techniques/processes to software development, the nuclear industry is probably 10-20 years behind DoD. Unlike DoD, most of the software these people develop is narrow in scope and much smaller in size. The people who write the code are not programmers in many cases and typically use FORTRAN. It wasn't all that clear that many were interested in expert systems. It was clear that many did not like working with programmers and therefore usually didn't. It also seemed that many of them (especially the managers) are clearly overwhelmed with the volume of techniques and processes involved with software. Unlike DoD, the NRC does not mandate much of anything with regard to how software should be developed (languages and processes). They do dictate things such as the kinds of systems and information that must be available via computer in a nuclear power plant.

Interesting comments were received from John Rushby of SRI concerning both the survey and ES verification and validation. He thought the survey work was informative and valuable. He also indicated that much of their work is shifting away from expert systems (they have only 1 project left). This seemed to be due, in part, to the prevalent industry mentality that ES is some magical paradigm that does everything for you. He made one interesting observation that error-proof systems are impossible and that it would be more cost effective to show that your program does not do anything harmful than to show it satisfies a spec.

In summary, the conference was worthwhile. In some respects, it was longer than it needed to be since the papers tended to overlap a lot in V&V content and the same issues seemed to be raised and re-raised. Despite this, the conference clearly indicated that development of guidelines is the right thing to do.

Copies of this publication have been deposited with the Texas State Library in compliance with the State Depository Law.